Journal of Protein Chemistry, Vol. 16, No. 6, 1997

Predicting Differential Antigen–Antibody Contact Regions Based on Solvent Accessibility

Frank J. Lebeda[1,2] and Mark A. Olson[1]

Received February 2, 1997

A novel computational approach was examined for predicting epitopes from primary structures of the seven immunologically distinct botulinum neurotoxins (BoNT/A-G) and tetanus toxin (TeTX). An artificial neural network [Rost and Sander (1994), *Proteins* **20,** 216] was used to estimate residue solvent accessibilities in multiple aligned sequences. A similar network trained to predict secondary structures was also used to examine this protein family, whose tertiary fold is presently unknown. The algorithm was validated by showing that it was 80% accurate in determining the secondary structure of avian egg-white lysozyme and that it correctly identified highly solvent-exposed residues that correspond to the major contact regions of lysozyme–antibody cocrystals. When sequences of the heavy (H) chains of TeTX and BoNT/A–G were analyzed, this algorithm predicted that the *most* highly exposed regions were clustered at the sequentially nonconserved N- and C-termini [Lebeda and Olson (1994), *Proteins* **20,** 293]. The secondary structures and the remaining highly solvent-accessible regions were, in contrast, predicted to be conserved. In experiments reported by others, H-chain fragments that induced immunological protection against BoNT/A overlap with these predicted most highly exposed regions. It is also known that the C-terminal halves of the TeTX and BoNT/A H-chains interfere with holotoxin binding to ectoacceptors on nerve endings. Thus, the present results provide a theoretical framework for predicting the sites that could assist in the development of genetically engineered vaccines and that could interact with neurally located toxin ectoacceptors. Finally, because the *most* highly solvent-exposed regions were not well conserved, it is hypothesized that nonconserved, potential contact sites partially account for the existence of different dominant binding regions for type-specific neutralizing antibodies.

KEY WORDS: Botulinum neurotoxin; neural network algorithm; neutralizing determinant; type-specific antibody; vaccine.

1. INTRODUCTION

A fundamental question in immunology concerns the underlying differences among structurally similar proteins with non-cross-reacting epitopes. Immunologically distinct members of homologous proteins are found in viral receptor glycoproteins (*e.g.*, seven serotypes of gp120 in human immunodeficiency viruses [VanCott *et al.*, 1994]) and in bacterial toxins (group A streptococcal pyrogenic exotoxin serotypes A–C [Gallis, 1984], and clostridial neurotoxins: tetanus toxin [TeTX] and seven botulinum neurotoxin [BoNT/A–G] serotypes [Schiavo *et al.*, 1994]). A previous study (Lebeda and Olson, 1994) noted that the C-termini of the clostridial neurotoxin heavy (H) chains

[1] Department of Cell Biology and Biochemistry, Toxinology Division, U.S. Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702-5011.
[2] To whom correspondence should be addressed.

0277-8033/97/0800-0607$12.50/0

contain a region of ~150 sequentially nonconserved residues. This rather long, sequentially hypervariable region led to the hypothesis that this segment is associated with the divergence in binding-type-specific neutralizing antibodies.

One test of this hypothesis is to identify the residues that contact these antibodies and determine whether they can be used to distinguish the serotypes. Identifying the amino acid residues that form contacts between proteinaceous antigens and antibodies, or protein ligands and their physiological receptors at complementary surfaces, is accomplished most precisely by examining the X-ray crystallographic structures of the complexes. This task, however, is so arduous and time-consuming that refining predictive techniques by using only primary structural information is an obvious necessity in furthering our knowledge of antigen and receptor epitope structures, and in enhancing vaccine or drug development.

Predicting antigenic regions from a protein's tertiary structure by using a residue's static accessibility to surrounding solvent molecules has been reasonably successful (Novotný and Auffray, 1984; Novotný *et al.*, 1987; see also Getzoff *et al.*, 1987; Geysen *et al.*, 1987*a*, *b*). Other analyses using three-dimensional (3D) information have also used C_α-backbone angle flexibility and electrostatic interactions (Thornton *et al.*, 1986; Getzoff *et al.*, 1987). Several approaches that have been in the public domain for a number of years analyze single protein sequences (e.g., Jameson and Wolf, 1988). A more recently developed neural network algorithm, originally designed for secondary structural predictions, was modified to estimate a residue's level of solvent exposure based on a training set composed of residue accessibilities calculated from known crystal structures (Rost and Sander, 1994*b*). With the exception of the program developed by Novotný and Auffray (1984), few of the previous approaches used this adaptive algorithm's ability to analyze multiple alignments to improve their predictions. Moreover, assessment of the results from previous studies (e.g., Thornton *et al.*, 1986) was limited by the lack of immunogen–antibody or ligand–receptor structural data and an accurate identification of contact residues.

The present study used this network algorithm to characterize the solvent accessibility of neurotoxin H-chain regions by static accessibility as an index of protein–protein contact. The best available information regarding these interactions (provided by five different avian egg-white lysozyme–antibody cocrystals) was exploited here to assess the ability of this algorithm to locate potential contact residues. Using this analysis as a guide, the solvent-exposed residues were predicted for the heavy (H) chains of a family of clostridial neurotoxins whose 3D folds are presently unknown but for which some experimental evidence of antigenic regions exists. Thus, the present study extends previous results on these neurotoxins (Lebeda and Olson, 1994) and compares these predictions with relevant experimental data.

2. METHODS

PHD, a neural-network algorithm, was used to make independent predictions on protein solvent accessibility and secondary structure starting with amino acid sequences. Input files to PHD contained single-amino acid sequences of hen egg-white lysozyme (HEL; Brookhaven Protein Data bank file 1HEL) or the various neurotoxin serotypes. For lysozyme, retrieval and alignment of 67 similar sequences from the SwissProt database was automatically performed by the program HSSP (Sander and Schneider, 1991). The amino acid sequences of the neurotoxin H-chains (TeTX, BoNT/A–G) that were used in this study were described and analyzed in detail previously (Lebeda and Olson, 1994). Sequences of three isoforms of the F serotype were also examined. In addition to analyzing single sequences, multiple alignments of their primary structures were made with MACAW (Schuler *et al.*, 1991) or with HSSP. As with HEL, after reading an input file containing a single sequence, HSSP retrieved and aligned other H-chain sequences from the database. A comparison of the outputs from these two alignment programs revealed that while MACAW inserted gaps, HSSP also deleted short segments to optimize the final alignment.

Output files from PHD contained a section that assigned a 10-state solvent accessibility value (range 0–9) per residue (PHD version 594–317 [Rost and Sander, 1994*b*]). The most highly solvent-exposed residues were operationally defined as having accessibility states 8 or 9; the remaining highly solvent-exposed residues were operationally defined as having an accessibility state of 7 with a reliability index of ≥4. Individual values

were sorted into three-state accessibility bins in a manner similar to the procedure described by Rost and Sander (1994*b*).

In another section of the output, PHD provided independently determined secondary structural predictions and assigned the conformations of all amino acids to one of three states: helix (α, 3_{10}, π), strand (extended-strand state), or other (the remaining, nonperiodic structures: bend, turn, coil, β-bridge) (PHD version 494–317 [Rost and Sander, 1993; Rost *et al.*, 1994*a*]).

The last ~150 residues within each of the C-terminal regions examined (BoNT/A–F and TeTX) were aligned according to the secondary structural elements predicted by PHD. The distribution of residues (based on their hydropathic properties) predicted to be in highly solvent-exposed, loop regions was quantified by calculating joint probabilities of occurrence. The hydropathies of amino acid side chains were classified by subdividing the consensus hydrophobicity scale developed by Eisenberg *et al.* (1984) into polar (D, E, H, K, N, Q, R), moderately polar (intermediate; C, G, P, S, T, Y), and hydrophobic (A, F, I, L, M, V, W) residues. Histograms of these probabilities for each residue location were constructed along with superimposed smoothed curves derived from a moving average calculation (window size = three locations).

To validate the use of highly solvent-exposed residues as a measure of detecting protein–protein contract residues, published crystallographic data on identified contact residues in avian lysozyme–antibody cocrystals were used. Contact residues in those studies were based on measured distances (typically <3.5 Å) required for H-bonding and significant van der Waals interactions. Four different HEL–antibody cocrystal complexes were used: HyHEL-5 (Sheriff *et al.*, 1987), HyHEL-10 (Padlan *et al.*, 1989), D1.3 (Fischmann *et al.*, 1991), and D44.1 (Branden *et al.*, 1994). One pheasant egg-white lysozyme (PHL)-monoclonal antibody Fv fragment complex was also used in this study (D11.15 [Chitarra *et al.*, 1993]). The HEL/PHL contact residues identified within these five complexes are summarized in Table I. The single species difference in these contact residues was N113 in HEL and K113 in PHL.

The static accessible surfaces of the HEL protein residues were determined with DSSP (Kabsch and Sander, 1983), which calculates the path of a simulated water molecule (spherical radius = 1.4 Å) along a protein's van der Waals surface (Novotný *et al.*, 1986). Secondary structure and residue accessibility were calculated by the DSSP option contained in the IDEAS program package (Kanehisa, 1988). Accessibility values from the HEL data set were smoothed by using a four-residue moving-average window calculation.

The antigenic index algorithm of Jameson and Wolf (1988) (which incorporates standard algorithms such as Kyte–Doolittle for hydrophilicity, and Chou–Fasman and Garnier–Osguthorpe–Robson for secondary structures) for single sequences were used within the GCG package (Devereux *et al.*, 1984) (Version 7.2, Genetics Computer Group, Madison, WI, 1991). Outputs for the HEL analysis were sorted into 10-state bins, and only the results from the three highest sets of values that corresponded to HEL contact residues were plotted (Fig. 1); for clarity, the few false-positive points are not depicted. All of the outputs from this analysis of the BoNT/A H-chain were used in Fig. 3.

Table I. Antigen Contact Residues from Five Avian Lysozyme–Antibody Cocrystals Used in This Study

HEL cocrystal	Resolution (Å)	Refinement (*R* value)	Lysozyme residues and locations	References
HyHEL-5	3.0	0.245	Q41, T43, N44, R45, N46, T47, D48, G49, Y53, G67, R68, T69, P70, L84	Sheriff *et al.* (1987)
HyHEL-10	3.0	0.24	H15, G16, Y20, R21, W63, R73, N74, T89, N93, K96, I98, S100, D101, G102	Padlan *et al.* (1989)
D1.3	2.5	0.184	D18, N19, G22, S24, N27, G117, T118, D119, V120, Q121, I124, R125	Fischmann *et al.* (1991)
D44.1	2.4	0.25	Q41, R45, T47, G49, Y53, R68, S81	Braden *et al.* (1994)
D11.15-PHL	2.5	0.216	R21, Y23, N103, N106, R112, K113, K116, G117, D119	Chitarra *et al.* (1993)

It should be emphasized that the five antibodies present in the cocrystals were all fragments of monoclonal antibodies. The binding affinities of these fragments and their contact sites on the antigen probably differ when compared to those associated with the polyclonal antibodies that are typically produced in immune responses (Atassi, 1978). In this report no assumptions were made regarding binding affinities. Nevertheless, in order to assess the model predictions quantitatively, the most complete available data set was used in which the contact residues were identified directly by X-ray crystallographic analysis of these antigen–antibody complexes.

3. RESULTS

3.1. Verification of the PHD Neural-Network Algorithm for Solvent Exposure: Prediction of Antibody Contact Regions in Avian Lysozyme

The solvent accessibility calculations from DSSP, which used experimental data from the crystal structure of HEL were initially compared to the results from the two different prediction algorithms. The middle solid waveform (Fig. 1, top) was smoothed by using data derived from the original DSSP analysis (superimposed dashed curve; see Section 2) and shows the five regions in HEL of highest solvent accessibility. The filled circles (below these curves) locate the positions of lysozyme residues (Table I) that were shown by crystallographic measurements to make contact with five different antibodies. The triangles and squares connected by vertical lines immediately below these circles indicate those residues which were predicted by PHD to be highly exposed (filled triangles = accessibility states 8, 9; filled squares = state 7). These predicted, highly solvent-exposed residues correspond to the five presently known antibody contact regions for this enzyme. The topmost solid curve (with inverted triangles) represents the output from the antigenic index analysis. This algorithm produced a pattern for antigenic binding sites that was comparable to PHD's prediction of highly solvent-exposed residues.

The pair of plots at the bottom of Fig. 1 show the general similarity between the DSSP analysis and the PHD prediction for the secondary structure of lysozyme. PHD was 80% correct on a per-residue basis for this three-state secondary structure prediction. As shown by the location of the identified HEL contact residues, the most highly solvent–exposed residues (states 8, 9) were predicted by PHD to occur mainly in nonstructured regions (i.e., loops, etc.).

3.2. Prediction of Highly Exposed Residues in Clostridial Neurotoxins

The sequence alignment presented in Fig. 2 shows that a long (~150 residue) segment of dissimilar amino acids was predicted to be present at the C-termini of all the neurotoxin H-chains. A summary of three different analyses is illustrated in Fig. 3. The upper graph illustrates the antigenic index analysis which only used the single sequence for BoNT/A. Most of the highly solvent-exposed residues (inverted triangles) occurred in the C-terminal half of the H-chain (the C-fragment; Fig. 3, residues 861–1296 [Middlebrook 1994, 1995]). The relative lack of exposed sites on the N-terminal half of the H-chain corresponds to the hydrophobic region (~80 residues) that was noted previously in the study of transmembrane ion channels, which are presumably formed by some of these residues (Lebeda *et al.*, 1994; Lebeda and Olson, 1995). Results consistent with this idea have also been obtained with BoNT/A, B, E and TeTX (Montal *et al.*, 1992; Ledoux *et al.*, 1994).

PHD produced results similar to those from the antigenic index when only single sequences were analyzed (data not shown). Different results were obtained from PHD, however, when eight sequences were aligned by using HSSP. In Fig. 3 (middle), the major portion of residues predicted to have the highest solvent accessibilities (states 8, 9) were clustered at the N- and C-termini of the H-chains (filled triangles = accessibility states 8, 9). The horizontal bars in Fig. 3 labeled H435–611 and H1150–1289 represent immunogenic peptide regions and will be discussed in the next section. The remaining highly accessible residues were located throughout the H-chain (filled squares = state 7). As noted above with the antigenic index output, there was a relative lack of highly solvent-exposed residues in the hydrophobic region of the N-terminal half of the H-chain.

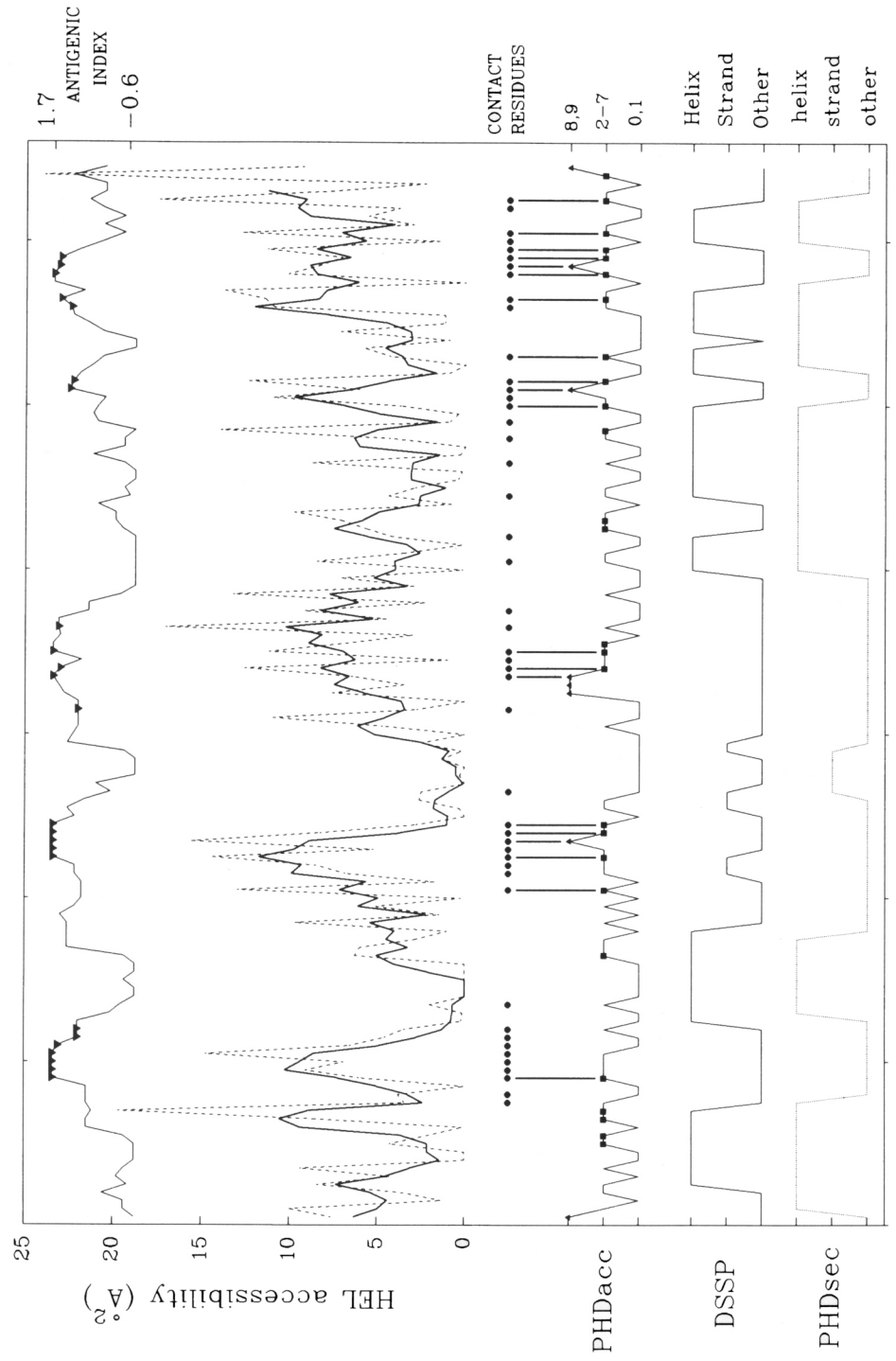

Fig. 1. Predictions of hen egg-white lysozyme (HEL) regions that contact different antibodies. Top: antigenic index prediction of epitope locations (solid curve) with locations of the three highest sets of calculated values (inverted triangles; see Section 2). Middle: original DSSP accessibility calculation from the HEL crystal atomic coordinates for single residues (dashed line), superimposed on the smoothed output from a four-residue moving average (solid curve) of the original DSSP data. Below these curves are the locations of the known lysozyme residues in contact with five antibodies (filled circles). Residues predicted by PHD to be the *most* highly exposed (states 8, 9) are represented by filled triangles. Residues predicted by PHD to be highly exposed (state 7; reliability index ≥ 4) are shown by the filled squares. Bottom: DSSP analysis of the secondary structure of HEL (helix, strand, or other; solid curve), and PHD three-state prediction of HEL's secondary structure (helix, strand, or other; dotted curve).

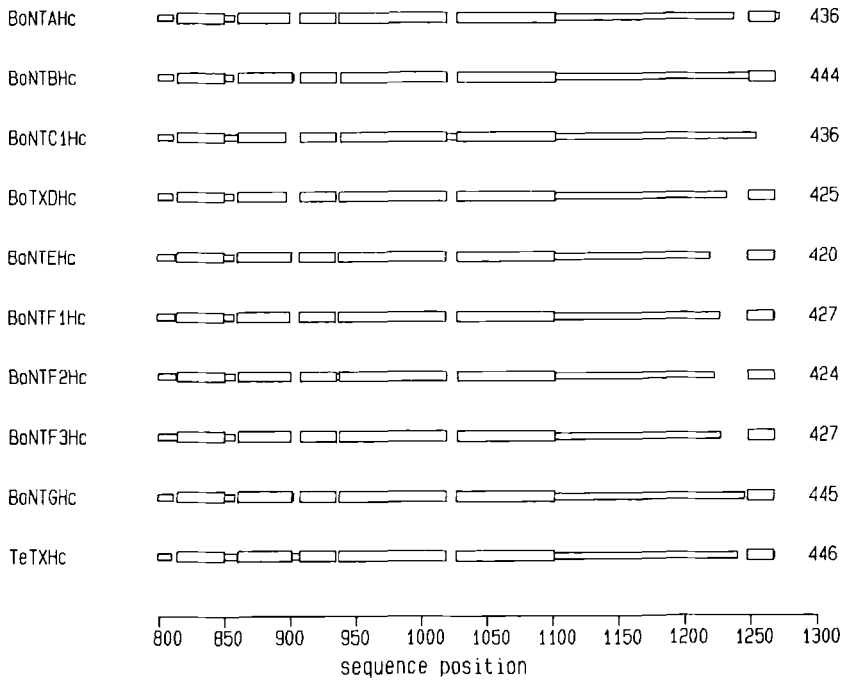

Fig. 2. Alignment of the neurotoxin H-chain primary structures with MACAW. Wide and thin bars represent sequentially similar and nonsimilar regions of residues, respectively. Long (~150 residues) nonsimilar regions occurred at the C-termini. Numbers on the abscissa represent the absolute locations of the BoNT/A H-chain residues. Numbers on the right indicate the number of residues within each C-fragment (Hc).

When aligned sequences obtained from MACAW were used as inputs, the independent determinations of secondary structures by PHD predicted that the C-terminal region of the neurotoxin H-chains was predominantly in a loop state (Lebeda and Olson, 1994). However, when HSSP was used to align these H-chains, more conserved secondary structural elements were predicted (Fig. 3, bottom plot; Fig. 4). The prediction of secondary structure by using the HSSP-derived alignment was considered to be more accurate than the prediction based on the MACAW-derived alignment because shorter loop regions were predicted using HSSP. These substantial differences in the secondary structural predictions are attributed to differences in the sequence alignments for the ~150-residue C-terminal segment in these neurotoxin H-chains (see Section 2).

An expanded view of the predicted, highly solvent-exposed region at the C-termini of the H-chains is presented in Fig. 4. The highly solvent-exposed regions (states 7–9) were aligned and were typically associated with conserved, nonperiodic secondary structures (loops, etc.). A similar pattern was noted for the entire C-fragment. Only three short regions comprised of four to six residues (vertical boxed areas) were predicted to be conserved and the most highly solvent-exposed (state 8 or 9).

The characteristics of the residues within these three short segments were examined further by calculating different joint probabilities of occurrence (Fig. 5). The three short regions that were noted in Fig. 4 to be locations for highly exposed residues are again represented in the box areas. It is clear from this joint probabilities analysis that there are several additional peaks that can now be distinguished along the C-terminal sequence that represent loops containing highly solvent-exposed residues (Fig. 5A). In more detailed analyses, the joint probabilities were calculated for highly solvent-exposed residues in loops having polar (Fig. 5B), intermediate polar (Fig. 5C), or hydrophobic (Fig. 5D) side chains. The majority of side chains were predicted to be either polar or moderately polar. Very few residues within these regions were

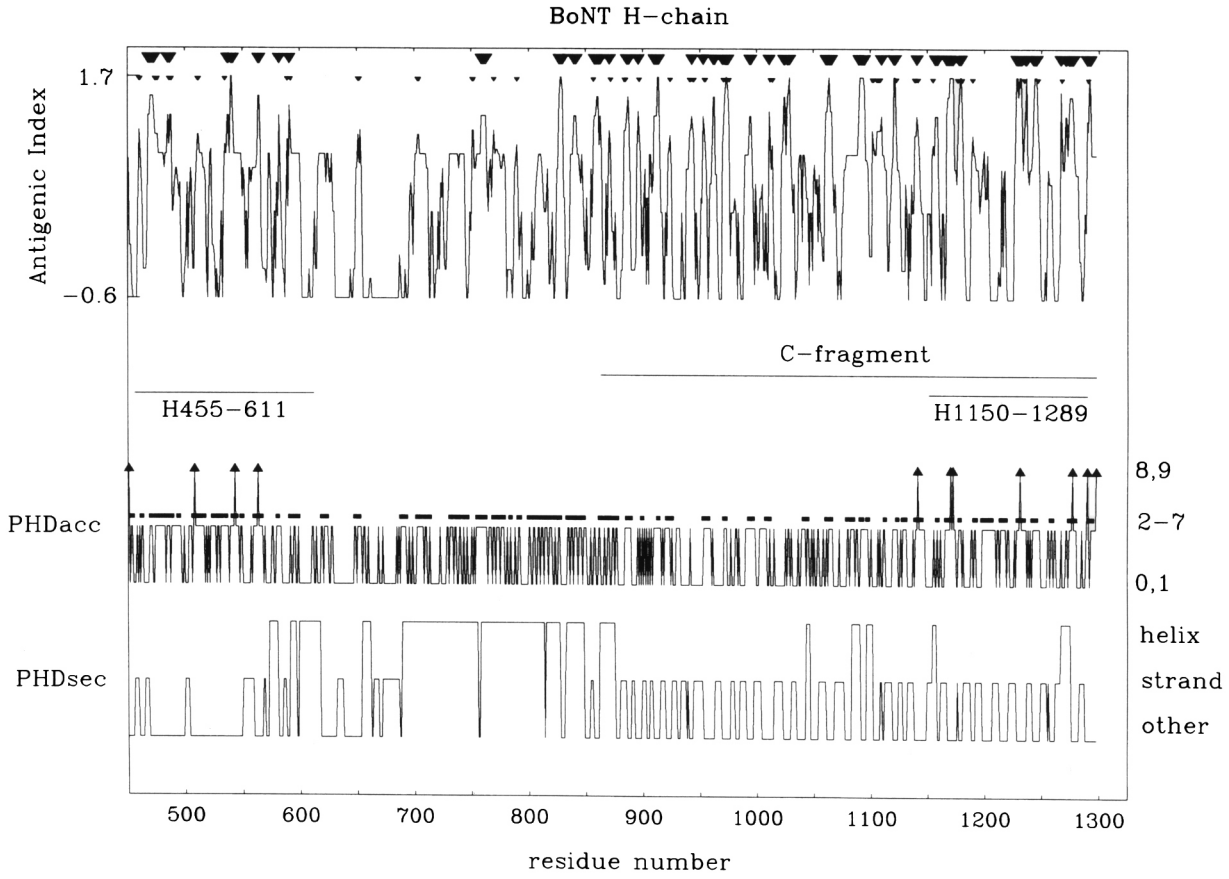

Fig. 3. Predictions of the antigenic index, residue solvent accessibility, and secondary structure of the H-chain of BoNT/A. Upper portion: ordinate values denote the antigenic index (using only the sequence of BoNT/A) that represents the degree of a residue being within an epitope. Locations having the two highest (large inverted triangles) and the third highest (small inverted triangles) sets of scores were clustered at the N-termini and overlapped with the C-fragment (horizontal bar). Middle portion: with a multiple aligned input with eight different serotypes, locations of residues are shown that were predicted by PHD to have high solvent-accessibility scores (states 8, 9: filled triangles; state 7: filled squares; reliability index ≥ 4). Compared to the results from the antigenic index, highly exposed residues were also predicted by PHD to be clustered at both termini. However, the locations that were predicted by PHD to be the *most* highly exposed were more restricted to the C-terminus. Lower portion: plot of three-state secondary structural predictions. Horizontal lines labeled H455–661 and H1150–1289 represent H-chain fragments that were individually capable of inducing protection *in vivo* (data from Dertzbaugh and West, 1995, 1996). Numbers on the abscissa represent the absolute residue locations on the H-chain of BoNT/A.

predicted to have hydrophobic side chains. Thus, from this probability analysis, there appear to be several locations within the C-termini of these neurotoxins that could be sufficiently exposed to make contact with other molecules such as antibodies or ectoacceptors.

4. DISCUSSION

Residue side-chain solvent exposure has been used previously as a predictive measure for identifying antigenic sites (Novotný *et al.*, 1986). Other researchers have proposed that protein antigen binding also involves C_α-backbone angle flexibility, electrostatic interactions, and rearrangements of the side chains in both the antibody and the antigen (Thorton *et al.*, 1986; Getzoff *et al.*, 1987). This study used a recently developed neural-network algorithm that was trained to estimate residue solvent accessibilities and to predict epitopic structures directly from sequences of amino acids. This technique was validated with avian egg-white lysozymes and five different antibodies (to date, the best characterized set of cocrystal data for a single antigen from the Brookhaven Protein Databank) by predicting the location of highly solvent-exposed regions that are in turn the most likely to make protein–protein

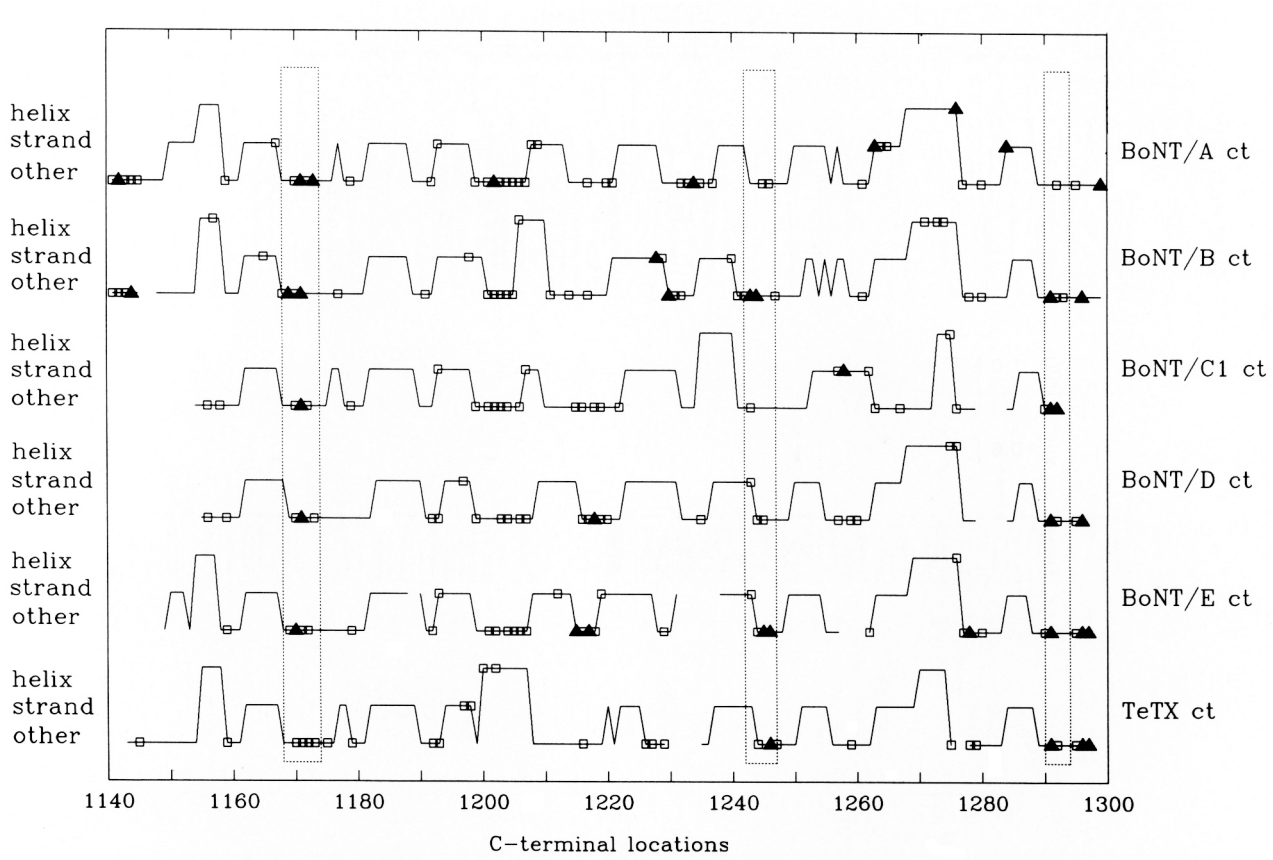

Fig. 4. Secondary structures and solvent accessibilities of neurotoxin C-terminal (ct) residues predicted by PHD from HSSP-aligned sequences. Although the primary structures of this region comprising ~150 residues were not conserved (Fig. 2), the secondary structural elements (helix, strand, other) and the highly solvent-accessible regions (open squares = state 7; reliability index ≥ 4), in general, were aligned. Except for three aligned regions (dotted boxes), the basic difference among these sequences was the presence of nonconserved locations of the *most* highly solvent-exposed residues (filled triangles = states 8, 9). Numbers on the abscissa represent the absolute residue locations on the BoNT/A C-terminus.

contacts. Side chains of the residues comprising these five epitopes account for approximately half of the total solvent-accessible area of HEL (6778 Å^2). The highly solvent-exposed residues (states 7–9) were predicted by PHD to be associated with all five contact regions. The PHD program, as expected, also predicted that most of the residues in these exposed regions were in loop conformations rather than in periodic structural elements.

Clostridial neurotoxins have global similarities of 85–90% in their secondary structures (Lebeda and Olson, 1994). In that study it was suggested that the regions of nonconserved amino acids in the H-chain would be important in defining the neutralizing epitopes. The longest sequentially nonconserved region appeared to be at the H-chain C-termini (Fig. 2 [Lebeda and Olson, 1994]). Despite the obvious differences in amino acid sequences, structural elements, along with the overall residue solvent accessibilities, were predicted to be conserved in these C-terminal regions. The important distinction among the serotypes at their C-termini was the general lack of conservation of regions that were predicted to be the most highly solvent-exposed (states 8, 9); only three regions were predicted to be conserved among the neurotoxin serotypes (vertical boxes, Fig. 4). As discussed below, these predictions are consistent with some of the available experimental observations.

The C-terminal halves of these toxins (C-fragments) contribute to the binding of these toxins to neural tissues. For example, the C-fragment of TeTX can decrease binding of the holotoxin in

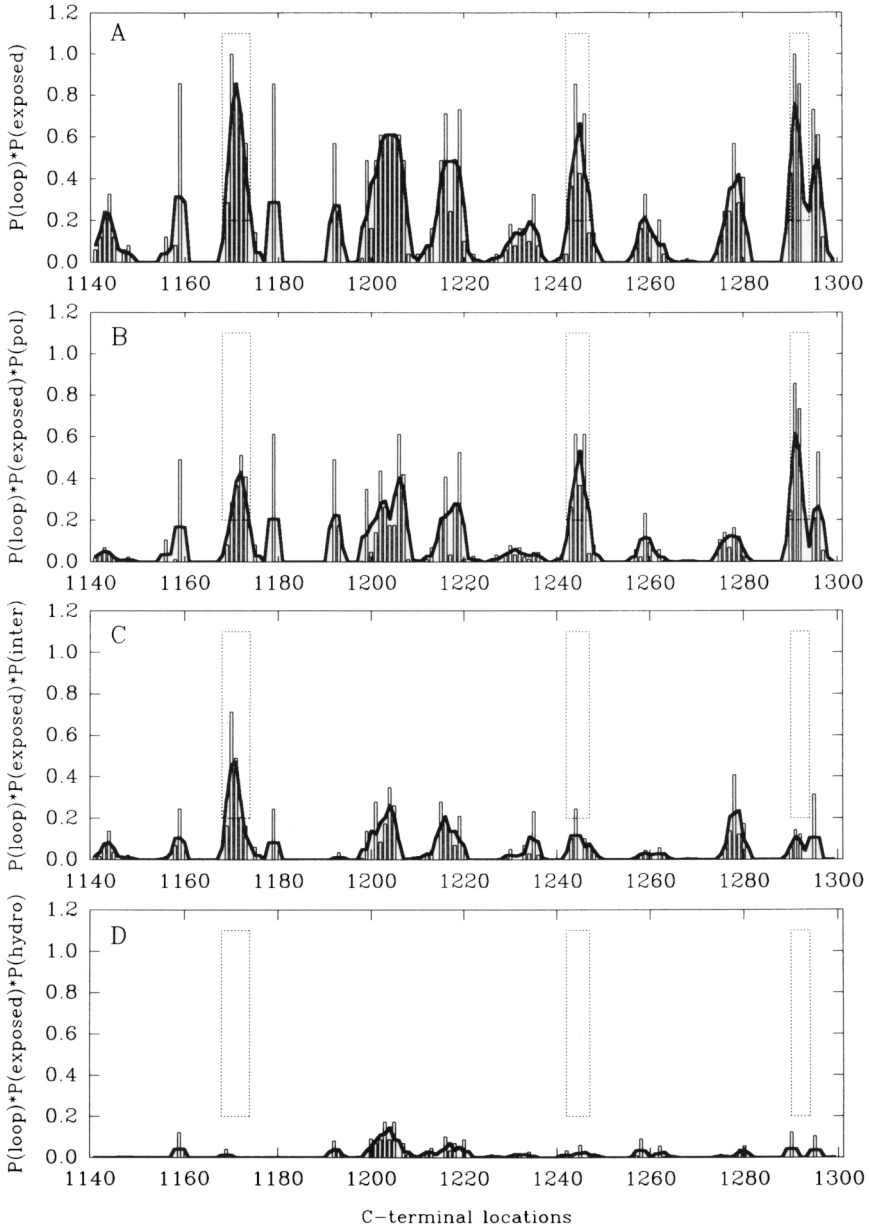

Fig. 5. Joint probabilities of occurrence for residue locations in the C-termini of seven different neurotoxin sequences. (A) The probabilities of a residue being highly solvent-exposed and within a loop. The three vertical boxes taken from Fig. 4 show the location of conserved, highly solvent-exposed residue locations. (B–D) The joint probabilities of locations having loops containing highly solvent-exposed residues that have polar, moderately polar (inter), and hydrophobic side chains, respectively. The dark solid curves superimposed on the histograms of the individual probabilities represent moving averages of three residue locations. Numbers on the abscissa represent the absolute residue locations on the BoNT/A C-terminus.

brain (Bizzini *et al.*, 1980). Although it has been well established that the A and the B serotypes of BoNT do not compete with one another for specific, high-affinity binding sites (toxin ectoacceptors) present at cholinergic nerve endings (Dolly *et al.*, 1987), it has been suggested from physiological experiments that the C-fragment of TeTX competes with the holotoxin for binding. Simpson (1984) also

suggested that the C-fragment of TeTX competes with the BoNT/C1 and BoNT/E holotoxins as well (but not with types A, B, D, F, or G). A possible noncovalent interaction between the C-fragment and the holotoxin, however, could not be ruled out. Nevertheless, it is possible that nerve endings may possess a single group or a number of similar toxin ectoacceptor molecules that have several binding sites for the different serotypes. In addition, there may be a conserved, highly solvent-exposed site on the H-chain C-terminus that has a recognition function (for the ectoacceptors or for antibodies) while the nonconserved, highly solvent-exposed sites take part in the remainder of the binding process. Furthermore, the presence of these nonconserved exposed regions supports the hypothesis that the C-terminal portion of the H-chain is an important region for binding non-cross-reacting neutralizing antibodies.

Second-generation vaccines developed from neurotoxin fragments to protect against clostridial neurotoxins are feasible, nontoxic alternatives to passive immunization with antitoxins and immunogenic toxoic products (Atassi and Young, 1985). Makoff *et al.* (1989) expressed efficiently the C-terminal half (~50 kDa) of the TeTX H-chain, which immunized and protected toxin-challenged animals. Unconjugated recombinant C-fragments of BoNT/A and /B induced a high degree of protective immunity against these holotoxins (Middlebrook, 1994, 1995). A combination of two unconjugated C-fragment-derived peptides also protected against lethal neurotoxin challenges (Smith *et al.*, 1995). In addition, a systematic analysis, using overlapping ~155-residue segments of the BoNT/A H-chain, showed that some protection occurred with an N-terminal and with a C-terminal fragment (Fig. 3, horizontal bars labeled H455–611 and H1150–1289, respectively [Dertzbaugh and West, 1995, 1996]).

The predicted regions of highly solvent-exposed residues on the toxin H-chains may be important for several reasons. First, the nonconserved, most highly solvent-exposed residues at the C-terminal may be critical for binding type-specific neutralizing antibodies and/or to ectoacceptors located on cholinergic nerve endings. Second, if these highly solvent-exposed residues are essential components of epitopes for neutralizing antibodies or toxin ectoacceptors, it is expected that replacement by one or more residues will alter binding affinities. Third, the correspondence between the theoretical predictions and the experimental evidence summarized in Fig. 3 (Middlebrook, 1994, 1995; Smith *et al.*, 1995; Dertzbaugh and West, 1996) suggests that the two clusters at either terminus of the BoNT/A H-chain contain highly solvent-exposed amino acids which may take part in a composite, spatially discontinuous structure that may characterize the dominant neutralizing epitope.

Finally, some of the predicted highly solvent-exposed regions at the N- and C-termini may be useful for developing shorter, more readily expressed antigens for future vaccine development. Overlapping peptides (19- to 22-mers derived from the C-fragment sequence) were recently examined for binding to anti-C-fragment antibodies (Atassi *et al.*, 1996). Those data are consistent with the predictions of this study in that highly exposed residues (states 7–9) are predicted to exist throughout most of the C-fragment. In addition, experiments are in progress (S. Bavari, personal communication) to determine whether binding occurs between neutralizing monoclonal antibodies from mice immunized with the C-fragment of BoNT/A and synthetic 25-mers whose sequences were derived from the predicted most highly exposed regions of that fragment.

ACKNOWLEDGMENTS

The authors thank Drs. Mark Dertzbaugh, John Middlebrook, Burkhard Rost, and Leonard Smith for their valuable comments and for sharing their results before publication. The continued support received from the staff at the Biomedical Supercomputing Center (Project 187001, National Cancer Institute, Frederick, Maryland) is also appreciated.

REFERENCES

Atassi, M. Z. (1978). Precise determination of the entire antigenic structure of lysozyme, *Immunochemistry* **15,** 909–936.

Atassi, M. Z., and Young, J. A. (1985). Discovery and implications of the immunogenicity of free small synthetic peptides, *CRC Crit. Rev. Immunol.* **5,** 387–409.

Atassi, M. Z., Dolimbek, B. Z., Hayakari, M., Middlebrook, J. L., Whitney, B., and Ishima, M. (1996). Mapping of the antibody-binding regions on botulinum neurotoxin H-chain domain 855–1296 with antitoxin antibodies from three host species, *J. Protein Chem.* **15,** 691–700.

Bizzini, B., Akert, K., Glicksman, M., and Grob, P. (1980).

Preparation of conjugates using two tetanus toxin derived fragments: Their binding to gangliosides and isolated synaptic membranes and their immunological properties, *Toxicon* **18,** 561–572.

Branden, B. C., Souchon, H., Eiselé, J.-L., Bentley, G. A., Bhat, T. N., Navaza, J., and Poljak, R. J. (1994). Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1, *J. Mol. Biol.* **243,** 767–781.

Chitarra, V., Alzari, P. M., Bentley, G. A., Bhat, T. N., Eiselé, J.-L., Houdusse, A., Lescar, J., Souchon, H., and Poljak, R. J. (1993). Three-dimensional structure of a heteroclitic antigen–antibody cross-reaction complex, *Proc. Natl. Acad. Sci. USA* **90,** 7711–7715.

Clayton, M. A., Clayton, J. M., Brown, D. R., and Middlebrook, J. L. (1995). Protective vaccination with a recombinant fragment of *Clostridium botulinum* neurotoxin serotype A expressed from a synthetic gene in *Escherichia coli, Infect. Immunol.* **63,** 2738–2742.

Dertzbaugh, M. T., and West, M. W. (1995). Identification of protective domains of botulinum neurotoxin serotype A, in *First International Conference on the Molecular Genetics and Pathogenesis of the Clostridia,* p. 12.

Dertzbaugh, M. T., and West, M. W. (1996). Mapping of protective and cross-reactive domains of the type A neurotoxin of *Clostridium botulinum, Vaccine* **14,** 1538–1544.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX, *Nucleic Acids Res.* **12,** 387–395.

Dolly, J. O., Ashton, A. C., Evans, D. M., Richardson, P. J., Black, J. D., and Melling, J. (1987). Molecular action of botulinum neurotoxins: Role of acceptors in targetting to cholinergic nerves and in the inhibition of release of several transmitters, in *Cellular and Molecular Basis of Cholinergic Function* (Dowdall, M. J., ed.), Ellis Horwood, Chichester, England, pp. 517–533.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with hydrophobic moment plot, *J. Mol. Biol.* **179,** 125–142.

Fischmann, T. O., Bentley, G. A., Bhat, T. N., Boulot, G., Mariuzza, R. A., Phillip, S. E. V., Tello, D., and Poljak, R. J. (1991). Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution, *J. Biol. Chem.* **266,** 12915–12920.

Gallis, H. A. (1984). Streptococcus, in *Zinsser Microbiology,* 18th ed. (Joklik, W. K., Willett, H. P., and Amos, D. B., eds.), Appleton-Century-Crofts, Norwalk, Connecticut, pp. 463–476.

Getzoff, E. D., Geysen, H. M., Rodda, S. J., Alexander, H., Tainer, J. A., and Lerner, R. A. (1987). Mechanisms of antibody binding to a protein, *Science* **235,** 1191–1196.

Geysen, H. M., Rodda, S. J., Mason, T. J., Tainer, J. A., Alexander, H., Getzoff, E., and Lerner, R. A. (1987*a*). Response to: Novotný *et al.*, 1987, *Science* **238,** 1584–1586.

Geysen, H. M., Tainer, J. A., Rodda, S. J., Masson, T. J., Alexander, H., Getzoff, E., and Lerner, R. A. (1987*b*). Chemistry of antibody binding to a protein, *Science* **235,** 1184–1190.

Jameson, B. A., and Wolf, H. (1988). The antigenic index: A novel algorithm for predicting antigenic determinants, *CABIOS* **4,** 181–186.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* **22,** 2577–2637.

Kanehisa, M. (1988). *IDEAS-88, Integrated Database and Extended Analysis System for Nucleic Acids and Proteins,* IDEAS User's Manual, Advanced Scientific Computing Laboratory, National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland.

Lebeda, F. J., and Olson, M. A. (1994). Secondary structural predictions for the clostridial neurotoxins, *Proteins: Struct. Funct. Gen.* **20,** 293–300.

Lebeda, F. J., and Olson, M. A. (1995). Structural predictions of the channel-forming region of botulinum neurototoxin heavy chain, *Toxicon* **33,** 559–567.

Lebeda, F. J., Hack, D. C., and Gentry, M. K. (1994). Theoretical analyses of the functional regions of the heavy chain of botulinum neurotoxin, in *Therapy with Botulinum Toxin* (Jankovic, J., and Hallett, M., eds.), Marcel Dekker, New York, pp. 51–61.

Ledoux, D. N., Be, X.-H., and Singh, B. R. (1994). Quaternary structure of botulinum and tetanus neurotoxins as probed by chemical crosslinking and native gel electrophoresis, *Toxicon* **32,** 1095–1104.

Makoff, A. J., Ballantine, S. P., Smallwood, A. E., and Fairweather, N. F. (1989). Expression of tetanus toxin fragment C in *E. coli*: Its purification and potential use as a vaccine, *Biotechnology* **7,** 1043–1046.

Middlebrook, J. L. (1994). Development of a molecularly engineered vaccine for botulinum toxin, *Toxicon* **32,** 242.

Middlebrook, J. L. (1995). Development of a molecularly engineered vaccine for botulinum toxin, in *First International Conference on the Molecular Genetics and Pathogenesis of the Clostridia,* p. 16.

Montal, M. S., Blewitt, R., Tomich, J. M., and Montal, M. (1992). Identification of an ion channel-forming motif in the primary structure of tetanus and botulinum neurotoxins, *FEBS Lett.* **313,** 12–18.

Novotný, J., and Auffray, C. (1984). A program for prediction of protein secondary structure from nucleic sequence data: Application to histocompatibility antigens, *Nucleic Acids Res.* **12,** 243–255.

Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A., and Rose, G. D. (1986). Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains), *Proc. Natl. Acad. Sci. USA* **83,** 226–230.

Novotný, J., Bruccoleri, R. E., Carlson, W. D., Handschumacher, M., and Haber, E. (1987). Antigenicity of myohemerythrin, *Science* **238:**1584.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J., and Davies, D. R. (1989). Structure of an antibody–antigen complex: Crystal structure of the HyHEL-10 Fab–lysozyme complex, *Proc. Natl. Acad. Sci. USA* **86,** 5938–5942.

Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol.* **232,** 584–599.

Rost, B., and Sander, C. (1994*a*). Combining evolutionary information and neural networks to predict protein secondary structure, *Proteins: Struct. Funct. Gen.* **19,** 55–72.

Rost, B., and Sander. C. (1994*b*). Conservation and prediction of solvent accessibility in protein families, *Proteins: Struct. Funct. Gen.* **20,** 216–226.

Sander, C., and Schneider, R. (1991). Database of homology-derived structures and the structural meaning of sequence alignment, *Proteins: Struct. Funct. Gen.* **9,** 56–68.

Schiavo, G., Rossetto, O., and Montecucco, C. (1994). Clostridial neurotoxins as tools to investigate the molecular events of neurotransmitter release, *Semin. Cell Biol.* **5,** 221–229.

Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). A workbench for multiple alignment construction and analysis, *Proteins: Struct. Funct. Gen.* **9,** 180–190.

Sheriff, T., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C., and Davies, D. R. (1987).

Three-dimensional structure of an antibody–antigen complex, *Proc. Natl. Acad. Sci. USA* **84,** 8075–8079.

Simpson, L. L. (1984). The binding fragment from tetanus toxin antagonizes the neuromuscular blocking actions of botulinum toxin, *J. Pharmacol. Exp. Ther.* **229,** 182–187.

Smith, L. A., Zuber, M., and Hoover, T. A. (1995). The protective efficacy of recombinant truncated Hc from *Clostridium botulinum* neurotoxin A, in *First International Conference on the Molecular Genetics and Pathogenesis of the Clostridia,* p. 23.

Thornton, J. M., Edwards, M. S., Taylor, W. R., and Barlow, D. J. (1986). Location of 'continuous' antigenic determinants in the protruding regions of proteins, *EMBO J.* **5,** 409–413.

VanCott, T. C., Bethke, F. R., Artenstein, A. W., McCutchan, F. E., McNeil, J. G., Mascola, J. R., Redfield, R. R., and Birx, D. L. (1994). Serotyping international HIV-1 isolates by V3 peptides and whole gp120 proteins using BIAcore, *Methods: Companion Meth. Enzymol.* **6,** 188–198.